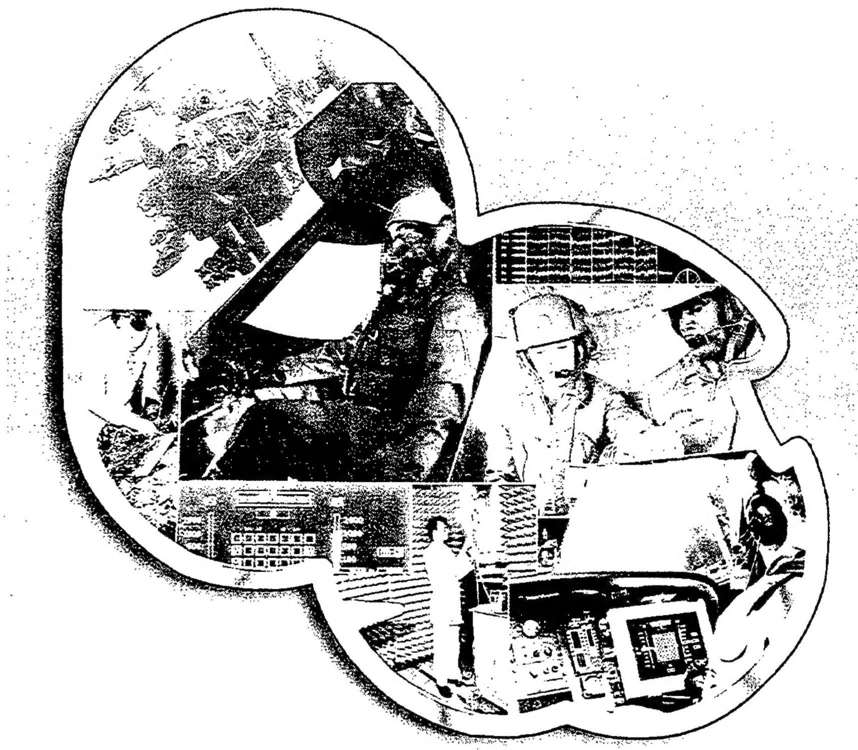
USAARL Report No. 2004-03

Nonspeech Audio in Helicopter Aviation

By Adrianus J. M. Houtsma



Aircrew Protection Division

December 2003

20040218 038

U S A A R L

U.S. Army
Aeromedical Research
Laboratory

Notice

Qualified requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), 8725 John J Kingman Road, Suite 0944, Fort Belvoir, Virginia 22060-6218. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about Laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 2003	Final

4. TITLE AND SUBTITLE

Nonspeech Audio in Helicopter Aviation

5. FUNDING NUMBERS

Program Element: 622787
Project Number: 878
Accession Number: DA 360347
Work Unit: 321

6. AUTHOR(S)

Adrianus J.M. Houtsma

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER

2004-03

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
504 Scott Street
Fort Detrick, MD 21702

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release, distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report describes the results of a literature study on the use of nonspeech audio in human-machine interaction, especially applied to helicopter aviation. The study focuses on various design philosophies for communicating important states of a machine to its operators through the use of nonspeech sounds. The two most popular sound designs are the so-called *auditory icons* (natural sounds that have a meaning by association with a real object) and *earcons* (abstracts sounds whose meaning must be learned). Usability indices such as effectiveness, efficiency and user satisfaction are reviewed for a number of experimental cases. The conclusion is that designs of auditory icons and earcons are not necessarily mutually exclusive. It appears entirely possible to develop sets of artificial sounds that combine the best features of each design approach.

14. SUBJECT TERMS

Nonspeech audio, auditory icons, earcons, helicopter navigation

15. NUMBER OF PAGES 17

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of contents

Page

Introduction

The human-machine interaction between the cockpit crew and a helicopter is typically multimodal and multisensory. A pilot's vision is heavily involved because the helicopter's surroundings must be continuously observed and instruments must be systematically monitored. Hearing is involved in processing verbal communication streams, both between aircrew and ground control and between individual crewmembers. Olfactory input may give the crew an early warning that somewhere something is burning. The vestibular system provides constant input about the body's acceleration in two directions of translation and three directions of rotation. Finally, the human motor control system reacts to all these sensory inputs by controlling the motion of the aircraft through mechanical manipulation of cyclic, collective and pedal controls. Dependent on the amount of direct mechanical coupling of these controls, they will not only react to the mechanical commands they receive from the pilot, but also provide mechanical feedback to the pilot through the human tactile, tactual and kinesthetic sense.

In this two-way multisensory interaction between man and machine, the largest workload appears to be assigned the visual system. Pilots typically receive a lot of 'visual training,' that is, they must learn to use their eyes in an effective and efficient manner by scanning surroundings and instruments in a systematic way. Because of the larger workload typically assigned to the visual system, it can easily become overloaded. In a combat situation, for instance, the visual system may find itself unable to keep complete track of rapidly changing and threatening surroundings, both outside and inside the aircraft.

Because we cannot point our ears the same way we can turn and focus our eyes, the two senses turn out to be complementary to a large extent. If, at a certain moment, our eyes are not turned and focused correctly, we can easily miss an observation that is essential to safety or survival. Our ears, however, are 'omnipresent.' They are always 'on' and ready to receive sound inputs from any direction. Through long-term practice in a same type of machine, pilots have become used to the typical sound patterns that are associated with different flight maneuvers. This sound is referred to in the literature as *functional sound*. Malfunctioning of the machine may often first be noticed by a change of perceived sound, before trouble is visually identified through the electronic warning system. Some aircraft have a limited set of electroacoustic warning signals for acutely dangerous situations.

The purpose of this report is to identify and evaluate possibilities for augmenting the use of the hearing sense in helicopter aviation, in order to alleviate the burden placed on the visual system and to lower overall workload and fatigue for the aircrew. The current situation with respect to the use of sound in aircraft is reviewed first. Then, an argument is developed for more use of sound communication in helicopter aviation. Next, a critical analysis of recent laboratory research on the effectiveness of two different types of electronic sound, *auditory icons* and *earcons*, is presented. Finally, recommendations are made for future research on the use of such sounds in typical helicopter environments, including the fact that many older pilots have suffered some degree of hearing damage.

Current situation

Helicopters are very noisy machines. Sound pressure levels of well over 100 dB inside the cockpit are quite normal. Because such sound levels are far beyond established hearing damage risk criteria (MIL-STD-1474, 1997) and occupational health standards (OSHA 1910.95, 1984), pilots must wear protective gear to prevent permanent hearing damage. The best protective gear available today is a combination of helmet-mounted earmuffs and the Communications Earplug (CEP), a pair of small telephones that are inserted in the ear canal and are sealed by an expanding foam tip (Mozo and Murphy, 1997). This earmuff-earphone combination not only provides double protection against environmental noise, but also provides clear electronic voice communication.

The noise and the functional sound, which a helicopter makes, however, are physically one and the same thing. This means that hearing protection could easily interfere with the perception of potentially vital acoustic warnings produced by the machine through its sound pattern. If the passive sound attenuation produced by helmet and CEP was frequency-independent, a deviant sound pattern would probably still be recognized by an experienced pilot. The actual sound attenuation, however, is much more efficient at high than at low frequencies, so that the total attenuated sound pattern is heavily tilted toward the low frequencies. A pilot who has learned through years of experience how a helicopter sounds while wearing a certain type of hearing protection, may have to do some relearning of those sound patterns when a new, more efficient type of hearing protection is applied. Although there are almost no hard data on these learning processes, informal observation of pilots' experiences seems to indicate that these adaptations are rather quick.

Currently, every helicopter type appears to have its own unique warning signal system. The UH-60 (Black Hawk), for instance, uses three very similar sounds that are meant to call the pilots' attention to a warning light display to identify the emergency situation. One sound is a low-frequency (about 300-Hz) harmonic complex tone that signals a too low rotor revolution speed. It cannot be turned off, and stops only when the low rotor rpm situation has been corrected. The very same sound is used to signal an engine failure (this one can be easily reset), while a periodically interrupted version of that same complex tone signals an imbalance between the stabilator actuators. The AH-64 (Apache) uses a very similar sound signal system. For the RAH-66 (Comanche) a much more complex audio warning system has been proposed comprising 36 pure-speech messages, 11 abstract sound signals, and 3 mixed sound/speech signals. The most urgent of these speech messages repeat the key word at the end, e.g., 'fire in left engine, fire,' to convey the urgency of the message and to provide some redundancy in case of noise interference. An important feature of warning signals is that they can be shut off once the arousal function has been achieved. Persistence of warning sounds can be very annoying and even dangerous (Patterson, 1989).

Why use sound?

The use of sound in addition to visual, vestibular, kinesthetic and olfactory sensory input during flight operations has several advantages. Human sound perception has a number of specific features that are not found in most other sensory systems. In principle, these features could be exploited to achieve more reliable communication and lessened workloads. We will discuss three of these features.

Omnipresence

Our eyes have eyelids that can be shut so that all visual perception is blocked. When our eyes are open and focused, sharp (foveal) vision is limited to an angle of only a few degrees (Hood and Finkelstein, 1986). Since our ears don't have 'ear-lids,' our hearing system is always open to sound, no matter its intensity, frequency content or direction. Although the hearing system is sensitive to direction of sound, there are no 'dead spots,' that is, there are no incident angles for which sound cannot be perceived. This makes a sound stimulus an excellent attention catcher since we don't have to make any specific effort to perceive it.

Independence

Although each of our sensory organs reveals specific aspects of our environment to our brain, these organs do not always operate independent of one another. The sense of smell is often tightly coupled to the sense of taste, for instance, when one tries to taste a piece of chocolate with the nostrils closed. Anatomically, our vestibular and auditory senses appear intertwined because they share a common peripheral structure in the temporal bone. Physiologically, however, the vestibular organ (sacculus, utriculus, semicircular canals) is coupled to the visual system via its neural connections, which one can easily verify by trying to focus on one's finger about two feet away with either the hand moving back and forth or the head turning left and right. One works, the other does not work at all. It is exactly this coupling that can cause a person to become disoriented or even sick when visual and vestibular inputs are made to contradict one another, which can easily happen on a ship or in an aircraft. Because the hearing system is rather independent of the visual and vestibular systems, at least at a peripheral level, it appears quite possible that a properly chosen auditory stimulus could resolve an apparent conflict between visual and vestibular perception, as happens in a case of spatial disorientation or vertigo.

Resolution power

Our visual system has excellent spatial resolution power, about 1 arc-minute in the foveal region (Olzak and Thomas, 1986), whereas the best auditory spatial resolution is about 2 degrees for sounds coming from straight ahead (Mills, 1958). In perceiving temporal variations of a stimulus, however, the ear outperforms the eye which is severely limited by a time constant of about 100 ms. The ear can 'follow' temporal variations in sound to a much higher degree,

resulting in sensations of rhythmic pattern for variations below 10 Hz, sensations of roughness for variations between about 10 and 100 Hz, and sensations of pitch for oscillations between 100 and several thousand Hz. Also, reaction times to auditory stimuli are generally shorter than for visual stimuli (Welch and Warren, 1986). The complementary nature of the eye's poor temporal and excellent spatial resolution versus the ear's excellent temporal and poor spatial resolution power may be used to divide potential input information up into visual and auditory parts, so that each sense gets to process the kind of information that it is optimally equipped for.

Despite some clear advantages of the use of sound in human-machine interaction, there also are some things to watch out for. Because we cannot turn off our ears as we can, to some extent do with our eyes, a continuous sound that has lost its informative function can quickly become annoying and even dangerous. Patterson (1989) describes several aviation accidents where, after an emergency sound alarm went off, the aircrew spent more effort trying to turn the sound off than on getting the aircraft under control. Alarm sounds should therefore be of sufficient intensity to alert the crew, but not so loud to annoy the crew or to interfere with intercrew communication. Alarm sounds also should be brief or easily resettable, since they quickly lose relevance after the crew has been alerted to an unusual situation. Simultaneous sound warnings also must be avoided because, without carefully designed sound signals or special 3D sound presentation equipment, simultaneous sounds will acoustically blend to a single unrecognizable and annoying nuisance. Finally, any sound signal to be used must have a specific meaning. Completely redundant sounds that carry no relevant information, as are found in many computer applications and electronic games, are to be avoided because they only lead to annoyance (Brewster, 2002).

Sound can have a variety of functions in aviation. These functions will generally fall into three main areas: communication, warnings, and navigation. *Communication* in the form of speech messages is the most common in aviation, both natural and electronic. Using newly developed technologies of canned or synthesized speech, it is now possible to let a machine verbally communicate with the crew. This is largely limited to simple one-way messages, however, since full two-way verbal interaction also requires automatic speech recognition, automatic language generation and human dialog simulation. Simple automatic message systems are now commonly found at many airports and subway systems. As *warning signals*, sounds generally are very effective because of their 'waking' power. A sound will readily tell a listener that there is a dangerous situation that needs attention, but information about what kind of danger and where to find it is another matter. This requires special sound design and listener training, about which more will be said in the following sections. For *navigation*, sound can potentially be used in a variety of ways. For instance, a frequently occurring situation where a pilot must keep a combination of horizontal and vertical needles centered in order to stay on a glide slope, could be replaced by a spatialized headphone-presented sound that must be kept in the center of the head. This not only alleviates the visual system, but may also result in quicker corrective stick action since auditory reaction times are considerably shorter than visual ones.

Types of potential sound signals

In this section, we will present three general types of sound which, one way or another, all have a potential role on aviation. *Speech sound* is the most natural and most frequently used type. On the various aspects of speech production and perception, there is a huge body of literature. Because this report is focused on nonspeech sound, this topic will only briefly be touched upon. *Auditory Icons* and *Earcons* are two different classes of artificial sound, each one designed to convey specific messages or meaning, but using different memory strategies.

Speech sounds

This class of sound contains all natural speech exchanged between crew members and ground control, either acoustically or electronically transmitted. Also, synthetic speech programmed into (parts of) the machine falls into this class. Advantages of using speech sound for communication, warning, or navigation is that the meaning is always clear as long as the language is understood, without the need of any extra learning process. Disadvantages of synthetic machine-generated speech are that it is long and becomes annoying when it is too redundant. There is evidence, however, that some degree of redundancy between simultaneous speech and visual signals decreases reaction time (Selcon, Taylor and McKenna, 1995). Parts of a speech message may also become masked by background noise, which could make a message unintelligible.

Auditory Icons

This term was first coined by Gaver (1989) to signify the use of synthetic sounds that have a natural intuitive meaning. They often are exact or very close imitations of sounds generated in the course of a natural physical process, for instance crumbling up a piece of paper and throwing it in a waste basket. Through our daily experience in the real world, we have been exposed to such sounds many times, and we have learned to associate sounds with specific processes. The idea of an auditory icon is that a sound, through its naturally learned association with a physical object or process, will trigger an appropriate meaning or interpretation when placed in a complex human-machine interaction task. This largely eliminates the need for training or learning process, which may be important if pilots have to deal with different aircraft types, each having its own nonspeech audio system. An example, taken from Gaver's 'Sonic Finder' developed for the Apple MacIntosh graphical user interface, is shown in Figure 1.

Figure 1. Deletion of folder in Gaver's (1989)
'Sonic Finder' Graphical User Interface.

When a folder is selected for deletion (top left), a 'papery' sound indicates that the target is a folder. The dragging of the folder toward the wastebasket (top right) is accompanied by a 'scraping' sound. When the pointer reaches the wastebasket (bottom left), a 'clinking' sound is heard. Finally, when the folder is dropped in the basket, it becomes 'fat' and a 'smashing' sound is heard.

The inherent strength of auditory icons is that they make use of natural long-term learned associations, and therefore require little or no training before use. Sounds are usually quite short, and interpretation is intuitive, fast and relatively unambiguous.

Since auditory icons are directly derived from natural sound processes, one may wonder to what degree sound features can be manipulated without losing the naturally-learned associations between sound and object. A more specific way to address this question is to find out to what extent specific properties of an object can be identified from merely hearing the associated sound. This line of questions has developed into an entirely new research field called 'ecological acoustics' (Gaver, 1993a, 1993b).

Several investigators have shown that people are able to identify object properties such as the lengths and materials of struck bars or the gender of a walking person from the sound produced by these processes. Houben (2002), who reviewed many of these studies, recently showed that properties like the size and speed of a rolling ball can be rather consistently identified from the associated sound. This implies that, if one is to use a rolling sound as auditory icon, one has the option of creating different versions of the rolling object, with different sizes, materials, velocities, etc. Especially if one has a good understanding of the sound source in the form of a quantitative physical model, one can create a large variety of other icons

within a same family (e.g. a rolling ball) corresponding to various velocities, sizes and materials, without running much of a risk that listeners may loose the natural intuitive association between sound and object. A physical model for the sound of a ball rolling over a suspended plate was recently developed by Stoelinga, Hermes, Hirschberg and Houtsma (2003).

An obvious limitation to the general use of auditory icons is the fact that it often is very difficult to find an appropriate sound for every function or process that takes place in human-machine interaction, especially when functions or processes are abstract and don't produce any natural sound. For example, what kind of sound should one choose for a 'stabilator actuator imbalance' warning in a helicopter? In such cases, one is inclined to select some arbitrary sound, and document this in the user manual so that it can become part of the pilot training. Such choices have, in fact, become known as *earcons*.

Finally, it should be noticed that an auditory icon does not always have to be a special sound, to be associated with an object or process. A process also can be supported acoustically by displaying a specific property of the process (e.g., spatial location) by means of sounds that already are there and are used in some other function. Imagine, for instance, a number of pilots flying in formation. The spatial position of each aircraft in the formation could, for each pilot, be superimposed on all speech communication signals by means of 3D sound display techniques. This way, each pilot hears the other pilots talk at spatial locations relative to their own head, congruent with the actual position of their respective aircrafts. It seems that this may be a powerful way to communicate one's position in a continuous manner, without having to cope with the addition of new and potentially annoying sounds.

Earcons

This term seems to have been coined by Blattner, Sumikawa, and Greenberg (1989) for abstract sounds, to accompany interface objects and processes that have a clear and recognizable hierarchical structure. The idea is borrowed from music. Sounds have no direct, intuitive or natural link with objects they represent, and all associations must therefore be learned. The hierarchical structure between and within classes of earcons used in an interface, however, is a direct reflection of the architecture and organization of the software system that is used.

The simplest form of earcons is the kind of sounds we typically hear when we open or close Microsoft Windows. They are short quasi-musical sounds that tell us whether or not, and when, a particular computer function is being activated. The sounds have no inherent or intuitive meaning. Sounds are often hard to document in a manual because of their abstract nature and the absence of an adequate formal musical alphabet to describe them (for instance, Windows' 'ta-tah' sound). Their meaning can only be learned through experience of running the software.

An example of a typical complex earcon system, taken from studies by Brewster, Raty, and Kortekangas (1995, 1996) is illustrated in Figure 2. The system begins with a simple steady sound at the top. Every time one travels down one node in the network, a specific musical feature is added to the sound, which remains in tact all the way down a branch. This way sound

images become more complex if one goes further down the network. By feature extraction of each sound, however, one can derive at which level and along which branch one is located. Figure 3 shows the organization chart of a particular software system. The basic idea of an earcon system is to design a sound system as shown in Figure 2 with the same nodal structure as the software system of Figure 3. Every node of the software then has a unique earcon. Figure 4 shows a typical training and test system in which a user, after hearing a sound (earcon), must identify the node that represents the current status of the software system.

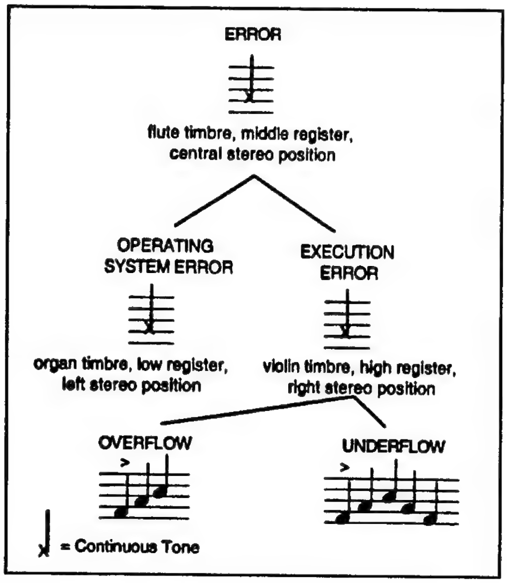


Figure 2. A hierarchy of earcons representing errors (from Blattner, Sumikawa, and Greenberg, 1989)

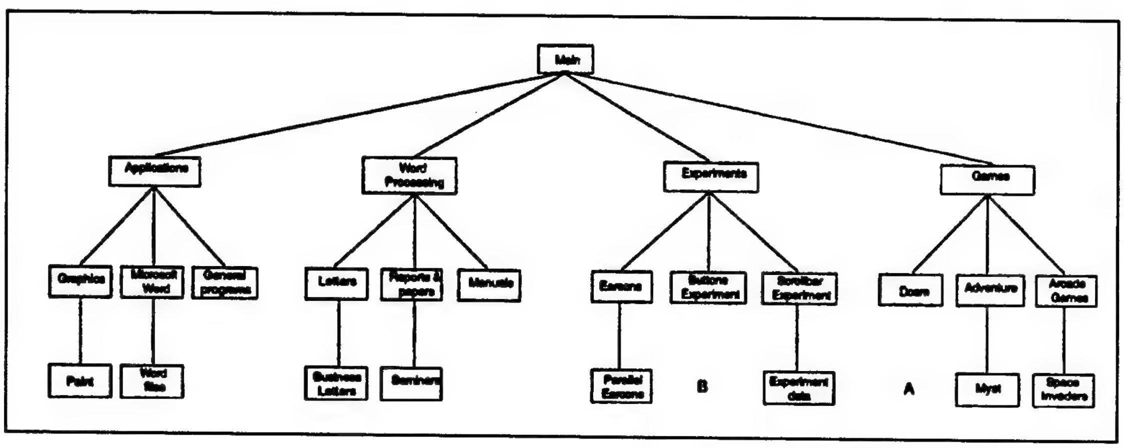


Figure 3. Diagram of a typical file system to be augmented with earcons.

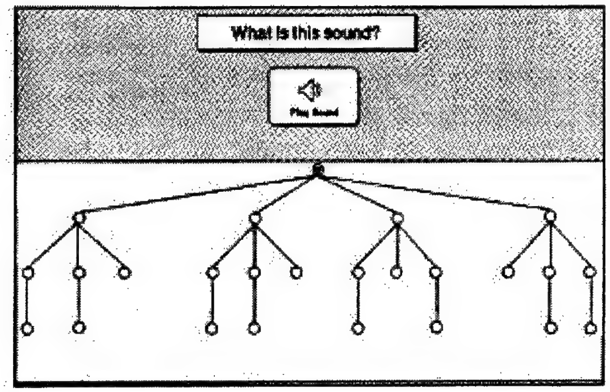


Figure 4. Training and test setup for learning to identify earcons.

The strong points of earcons appear to be the underlying hierarchy and logic of the earcon system. This logic is directly borrowed from classical Western music where similar hierarchical structures are used. Part of music appreciation is learning to recognize these structures, so that listening becomes an active process of exploration, anticipation and surprise.

Weak points of earcons are their abstract nature and, consequently, the absence of any direct association of sounds with real things. All associations have to be learned, with the support of the underlying logic of the system. This is time consuming and slow, and therefore, from the onset, seems unpractical for use in aviation where quick and unambiguous reactions are often required. Moreover, because of its inherent arbitrariness of design, an earcon system makes sense only if it was implemented in the same way by every designer. Otherwise, there would be too much learning and relearning. Standardization of earcon systems, however, seems, for the time being, more idealistic than realistic. After all, sign language for the deaf, which could in principle be universal and language-independent, is in fact a rather chaotic collection of many different sign systems across the world.

Performance evaluation of nonspeech audio

The introduction of auditory icons and earcons in graphical user interfaces during the early nineties was mostly presented as design work in the literature. The International Conference on Auditory Display (ICAD), started in 1992, and the annual Computer Human Interaction (CHI) conference appear to have been the principal platforms for publication and exchange of ideas. Most of the papers found in the proceedings can be characterized as design papers, merely describing creative ideas and implementations. Systematic evaluations of design usability, in terms of effectiveness, efficiency and user satisfaction, are mostly sketchy, if done at all, and appear very biased because designers evaluated their own designs. The one thing, however, that stands out among sparsely provided performance information is that graphical user interfaces, augmented with either speech, auditory icons or earcons, found ready acceptance among visually impaired and blind computer users (Brewster, 2002).

Earcons

Most of the performance evaluations of earcons seem to have been done by Brewster's group in Glasgow (Brewster, Wright, and Edwards, 1992; Brewster, Raty, and Kortekangas, 1995,1996; Brewster, Capriotti, and Hall, 1998). Summarizing results on recognition scores for earcons, they found that:

1. Navigation in a system network with 4 layers and a total of 25 nodes equipped with earcons at every node, the earcon recognition score was about 80%.

2. Reduced sound quality, as is typically found in mobile telephones, reduces this score to about 70%.

3. Compound earcons made of temporally concatenated sounds, can increase the recognition rate to almost 100%.

4. There is little or no difference between recognition performance of musicians compared with non-musicians.

A preliminary conclusion we can draw from this evaluation work is that recognition rates for earcons are well beyond the limits given by the 7±2 law (Miller, 1956), probably caused by the multidimensional nature of earcon stimuli. In assessing the performance, however, one must take into account that listeners performed the task in a very limited context, without distraction or heavy workloads. The evaluation reports provide little information on the amount of time it took the subjects to reach their ultimate performance level. Neither is there any information of the type of recognition errors that occurred. If earcon sounds are to be applied in aviation, it is important to know what kind of recognition errors are made once a sound is misinterpreted, since the ensuing action might be catastrophic. Hence, the overall conclusion seems to be that earcons are not the most appropriate types of sounds to apply in helicopters.

Auditory icons

Systematic performance studies of the effectiveness and efficiency of auditory icons are sketchy and sparse as well. Stevens (2002) compared learnability of auditory icons that were either 'ecological' (i.e., its sound shared all features with the real natural sound) or 'metaphorical' (i.e., the sound shared some features with the intended object's sound). She found that learning is about equally effective in the end, but 'ecological' icons are adopted more rapidly than 'metaphorical' icons. The overall recognition rates she obtained ranged from 75 to 95%.

Houben (2002) studied subjects' ability to discriminate between diameters and to absolutely identify velocities of rolling balls from their recorded sounds. He found that diameter discrimination strongly depended on the absolute size of the ball. Almost perfect scores were found when a 45-mm ball was to be distinguished from a 55-mm ball, both rolling at the same

speed (0.75 m/s). Discrimination between a 25-mm and a 35-mm ball, however, yielded scores of only about 75% correct. Absolute identification for six different velocities of one rolling ball by six subjects yielded a large range of scores, between 11% and 93% correct, with an average of 70%. The results indicate that the relationship between physical magnitude of an object's features and audibility of those features is quite complex. They also suggest that the perceptual process of feature extraction is subject to some degree of learning.

Comparison between Auditory Icons and Earcons

Perhaps the most comprehensive comparative performance evaluation of auditory icons and earcons was recently done in a series of studies performed at the Catholic University of Nijmegen (Bussemakers and de Haan, 2000; Lemmens, Bussemakers, and de Haan, 2000, 2001). Their general paradigm was to measure the total response time necessary to classify a suddenly presented picture in the presence or absence of a simultaneous earcon or auditory icon that could be either congruous or incongruous with the presented picture. Earcons were simple major or minor triad chords, respectively assigned to picture classes of, e.g., animals and non-animals. Auditory icons were animal sounds that either matched or did not match the animal shown in the picture. The absence of an icon or earcon (silence) was also an experimental condition. Pictures and sounds started synchronously on each trial, and response times were measured relative to that time. All sounds had durations of 1.22 seconds, i.e., responses are typically given during the period that pictures and sounds are on. For earcons, it was found that the shortest response time was achieved with silence, and the longest response time with incongruous earcons. For auditory icons, silence yielded the longest response times and congruous icons the shortest. Results are shown in Figure 5.

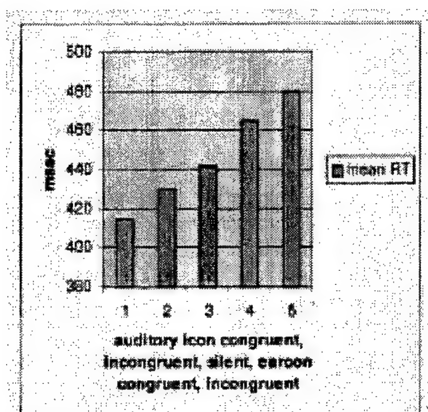


Figure 5. Classification response times for congruent and incongruent icons and earcons (from Bussemakers and de Haan, 2000).

Considering that these are choice reaction times, and that the basic overhead of reacting to any visual stimulus is about 250 ms, the differences in response time shown in Figure 5 are quite substantial. A reasonable overall conclusion from the findings of the Nijmegen group seems to be that, within the conditions of this experiment, earcons seem to have an inhibitory effect on

recognition and classification performance, whereas auditory icons appear to have a facilitating effect.

Exploring spatial properties of nonspeech sound

Most design and performance studies on earcons or auditory icons found in the literature are, from an acoustic viewpoint, monaural. In most cases it is assumed that, because the nonspeech sound is to support human-computer interaction, the sound will simply come from the speaker set of the computer system. During the last decade, however, a few studies have been published that seem to indicate that the inclusion of spatial sound properties can significantly enhance the effectiveness and efficiency of nonspeech audio.

Begault (1993) compared the efficiency of a 3D audio display with that of a mono (single earpiece) audio display in a visual target acquisition task during simulated flight. The result was an average 2.2 seconds faster acquisition time for the 3D audio system, although the total number of captured targets was about the same. Begault and Pittman (1996) later found an average advantage 0.5 seconds in visual target acquisition time for a 3D audio Traffic Alert and Collision Avoidance System (TACAS) compared with a standard visual/mono-audio TACAS.

Teas (1994) investigated subjects' ability to both identify and localize each of five different auditory icons having different temporal and spectral characteristics and variable spatial positions presented through a virtual 3D audio generator. He found that identification was almost perfect, but that localization performance was highly variable.

Bronkhorst, Veltman, and van Breda (1996) compared acquisition time for moving targets in simulated flight using a virtual 3D audio display, a birds-eye radar, or a standard visual tactical display. They found that both the 3D audio and the birds-eye radar displays reduced target acquisition times significantly compared with the standard tactical display, and that simultaneous use of 3D audio and birds-eye radar reduced these times even further.

Tan and Lerner (1996) measured the potential effectiveness of various types of warning sounds (siren, buzzer, repeated tone, voice messages) presented through one (sometimes two) of 12 loudspeakers mounted at different locations in a passenger car. The warning sounds were to signal not only the fact *that* there was a danger, but also *where* the threat was coming from. Subjects were instructed to respond to sounds by moving a joystick as quickly and as accurately as possible in the direction of the warning sound. Dependent variables were *reaction time* (time required to start joystick movement), *decision time* (time required to finish joystick movement), and *accuracy* (azimuth difference between loudspeaker and final joystick position). Variance analysis showed statistically significant effects of actual speaker location and type of warning sound on all dependent variables, but did not provide much insight in possible causal relations. The study as a whole raised some technical and methodological questions. The conclusion that 'subjects can locate warning sound direction within reasonable time' should therefore be considered as tentative.

Recently, Johnson and Dell (2003) presented a critical review of airlines' experience with audio warning signals in Boeing passenger jets and the USAF's experience with warning signals in the F-16 fighter. The many problems reported by pilots prompted the investigators to perform a simulator-based experiment comparing the effectiveness of monaural (same sound in both ears), stereo (sound in just one ear), and virtual 3D sound with respect to reaction time to warning messages and to subjectively experienced workload. The results were disappointing in the sense that 3D spatialized sound was not more effective than sound in just one ear (stereo), and that either 3D or stereo sound were only marginally more effective than conventional monaural sound.

On the whole, results of using spatial properties of nonspeech audio appear sufficiently promising to warrant further exploration. The technique can add a powerful dimension to a sound, making it more functional. The 3D-sound display technique is also relatively inexpensive for the end user, since all that is needed is a pair of stereo headphones.

Research issues to be pursued

Hearing-impaired users

A very important issue that so far has hardly been explored and definitely needs to be pursued concerns the usefulness of nonspeech audio for users with impaired hearing. Since many of our helicopter pilots, particularly the generation that used to fly without adequate hearing protection, have acquired at least some degree of noise induced hearing loss, we must get a better understanding of the constraints and limitations that hearing loss imposes on the effectiveness and efficiency of auditory warning, communication, or navigation signals. An absolutely essential requirement is *audibility*. This implies not only that a sound is detectable, but also that it has sufficient arousal power to draw the user's attention and that all its temporal and spectral details are well within the user's perception range. For hearing impaired users, this may imply that sounds are acoustically well separated from the noise background of the cockpit, for instance, through use of passive and active hearing protection in combination with a voice/sound-transmitting earplug. A second important requirement is *intelligibility*, since voice messages and nonspeech sounds must be correctly interpreted. A third somewhat related requirement is *localizability*, if spatialization of sound is used as part of a transmitted message. For instance, simultaneous conversations over one single frequency channel, which are unintelligible with monaural audio, can be made quite intelligible by proper 3D spatial positioning techniques. Recent research (Lorenzi, Gatehouse, and Lever, 1999; Noble, Byrne, and Lepage, 1994) appears to imply that such techniques also may work for listeners with noise-induced hearing loss, provided that conversations are spatially positioned in the horizontal (azimuth) and not in the vertical (meridian) plane. The reason is that induced hearing loss appears to leave interaural time delay sensitivity largely intact, which provides the principal cue for a sound's azimuth position. Sensitivity for high-frequency spectral profile, which is the

principal cue for elevation of a sound source, is often severely diminished in this type of hearing impairment.

Spatialization of sounds

In general, spatial properties of sound should be explored and applied to a much greater degree in human-machine interfaces. Spatial location of sound is a very fundamental and direct percept that has always been essential for survival of most species. Therefore, adding the spatial dimension to communication, warning or navigation sounds is directly compatible with the very nature of our hearing system, and is entirely in line with the design philosophy of auditory icons. Experimental evidence so far seems sufficiently encouraging to continue these efforts.

Methods of introducing a spatial dimension without actually adding new sounds to the system should also be further explored. By spatializing ongoing conversations that occur between crewmembers, between pilots of different aircraft, or between pilots and central control, one can implement powerful auditory icon functionality without making pilots learn a new library of sounds.

Merging auditory icons and earcons

Auditory icons and earcons were so far presented as two distinct and different types of non-speech sound. Actually they represent two extremes of a continuous scale (Brewster, 2002). Starting from 'ecological' auditory icons, which are direct imitations of natural sound processes, one can systematically alter acoustic features on the basis of source models to derive new ecological icons that correspond to other natural sound objects of, for instance, different size, shape or material. One can, however, also change acoustic features in directions that exceed correspondence to natural sound objects. Stevens' (2002) 'metaphorical' auditory icons are an example of such an operation. Such a change of apparent object features beyond the 'natural' eventually must lead to completely abstract sounds that do not evoke direct object associations, which are in essence earcons. From a perceptual viewpoint, a lot more research is waiting to be done to explore how far such metaphorical transformations can be stretched before associations are lost, and how this affects learnability.

From a purely logical viewpoint it seems that it should be possible to combine the 'best features' of auditory icons and earcons. The strength of auditory icons is their direct cognitive associations and their minimal learning requirements. The strength of earcons is the underlying hierarchy and logic of a well-designed earcon system. There seems to be no way to attach associations to earcons without going through some learning phase. There is no apparent reason, however, why an auditory icon system could not have a logical and hierarchical structure. Selection of the right objects that already have a natural relationship, in combination with some systematic, model-based feature extension, should in principle enable the construction of a set of auditory icons that (1) do spontaneously evoke direct object associations, and (2) exhibit as a system a logical, hierarchical and recognizable structure. This possibility also should be further explored.

References

Begault, D.R. 1993. Head-up auditory displays for traffic collision avoidance system advisories: a preliminary investigation. Human factors. 35: 707-717.

Begault, D.R. and Pittman, M.T. 1996. Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. International journal of aviation psychology. 6: 79-93.

Blattner, M, Sumikawa, D., and Greenberg, R. 1989. Earcons and icons: their structure and common design principles. Human computer interaction. 4/1: 11-44.

Brewster, S.A., Capriotti, A., and Hall, C. 1998. Using compound earcons to represent hierarchies. Human computer interaction letters. 1: 6-8.

Brewster, S.A., Raty, V-P, and Kortekangas, A. 1995. Representing complex hierarchies with earcons. Norway. ERCIM Technical Report 05/95R037.

Brewster, S.A., Raty, V-P, and Kortekangas, A. 1996. Earcons as a method of providing navigational cues in a menu hierarchy. Proceedings of the HCT96. 167-183. Springer Verlag.

Brewster, S.A., Wright, P.C., and Edwards, A.D. 1992. A detailed investigation into the effectiveness of earcons. Proceedings 1st international conference on auditory display. 471-498. C. Kramer, editor. Addison Wesley.

Brewster, S.A. 2002. Non-speech audio output. The human computer interaction handbook. J. Jacko and A. Sears, editors. Lawrence Earlbaum Associates, Chapter 12.

Bronkhorst, A.W., Veltman, J.A., and van Breda, L. 1996. Application of a three-dimensional auditory display in a flight task. Human factors. 38: 23-33.

Bussemakers, M.P., and de Haan, A. 2000. When it sounds like a duck and it looks like a dog: auditory icons vs. earcons in multimedia environments. Proceedings of the international conference on auditory display 2000, 184-189. Atlanta.

Department of Defense. 1997. Noise limits. Washington DC. MIL-STD-1474.

Department of Labor. 1984. Occupational noise exposure. Washington DC. OSHA 1910.95.

Gaver, W.W. 1993a. What in the world do we hear? An ecological approach to auditory event perception. Ecological psychology. 5: 1-29.

Gaver, W.W. 1993b. How do we hear in the world? Explorations of ecological acoustics. Ecological psychology. 5: 285-313.

Gaver, W.W. 1989. The SonicFinder: an interface that uses auditory icons. Human computer interaction. 4/1: 67-94.

Hood, D.C. and Finkelstein, M.A. 1986. Sensitivity to light. Handbook of perception and human performance. K.R. Boff, L. Kaufman and J.P. Thomas editors. J. Wiley, Chapter 5.

Houben, M. 2002. The sound of rolling objects: perception of size and speed. Ph.D. thesis, Technical University Eindhoven, The Netherlands.

Johnson, C.W. and Dell, W. 2003. Limitations of 3D audio to improve auditory cues in aircraft cockpits. Proceedings of the 21st international safety system conference 2003. 990-999. Ottawa, Canada.

Lemmens, P.M.C., Bussemakers, M.P., and de Haan, A. 2000. The effect of earcons on reaction times and error rates in a dual-task vs. a single-task experiment. Proceedings of the international conference on auditory display 2000. 177-183. Atlanta.

Lemmens, P.M.C., Bussemakers, M.P., and de Haan, A. 2001. Effects of auditory icons and earcons on visual categorization: the bigger picture. Proceedings of the international conference on Auditory Display 2001. 117-125. Espoo, Finland.

Lorenzi, Ch., Gatehouse, S., and Lever, C. 1999. Sound localization in noise in hearing-impaired listeners. Journal of the acoustical society of america. 105: 3454-3463.

Miller, G.A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing Information. Psychology review. 63: 81-97.

Mills, A.W. 1958. On the minimum audible angle. Journal of the acoustical society of america. 30: 237-246.

Mozo, B.T. and Murphy, B.A. 1997. The assessment of sound attenuation and speech intelligibility of selected ANR devices and the Communications Earplug when used with the HGU-56/P aviator helmet. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory. USAARL Report 97-08.

Noble, W., Byrne, D., and Lepage, B. 1994. Effects on sound localization of configuration and type of hearing impairment. Journal of the acoustical society of america. 95: 992-1005.

Olzak, A. and Thomas, J.P. 1986. Seeing spatial patterns. Handbook of perception and Human Performance. K.R. Boff, L. Kaufman and J.P. Thomas editors. J. Wiley, Chapter 7.

Patterson, R.D. 1989. Guidelines for the design of auditory warning sounds. Proceedings of the institute of acoustics. 11/5: 17-24.

Selcon, S.J., Taylor, R.M. and McKenna, F.P. 1995. Integrating multiple information sources: Using redundancy in the design of warnings. Ergonomics. 38: 2362-2370.

Stevens, K. 2002. Recognition of auditory icons: the effect of signal-referent relation and manipulation of acoustic parameters. MARCS Auditory Laboratory, University of Western Sydney, Australia (seminar abstract).

Stoelinga, C.N.J., Hermes, D.J., Hirschberg, A., and Houtsma, A.J.M. 2003. Temporal aspects of rolling sounds: a smooth ball approaching the edge of a plate. Acta acustica. 89: 809-817.

Tan, A.K. and Lerner, N.D. 1996. Acoustic localization of in-vehicle crash avoidance warnings as a cue to hazard direction. Department of Transportation. Washington DC. Technical. Report DOT-HS-808-534.

Teas, D.C. 1994. The development of auditory icons for representation of virtual objects in 3-D space. U.S. Air Force Armstrong Laboratory, Wright Patterson AFB. Technical Report AL/CF-TR-1997-0102.

Welch, R.B and Warren, D.H. 1986. Intersensory interactions. Handbook of perception and human performance. K.R. Boff, L. Kaufman and J.P. Thomas editors. J. Wiley. Chapter 25.